Similaridade de Imagens pela Análise da Aleatoriedade Utilizando Compressão de Dados

Rafael Divino Ferreira Feitosa
Instituto Federal de Educação, Ciência e Tecnologia Goiano
Ceres, Goiás, Brasil
E-mail: rafael.feitosa@ifgoiano.edu.br

Anderson da Silva Soares
Instituto de Informática, Universidade Federal de Goiás
Goiânia, Goiás, Brasil
E-mail: anderson@inf.ufg.br

***Abstract*—Classical techniques of image pattern recognition are dependent on the extraction and features selection steps. However, in most of the problems the discriminant features are unknown. It is proposed a classifier of similar images, with pixel-based features and without of the selection step, using data compression. The results show that randomness can be used as discrimination measure, yelding 0.83 of accuracy in the test set.**

***Resumo*—Técnicas clássicas de reconhecimento de padrões em imagens são dependentes das etapas de extração e seleção de características. Entretanto, em grande parte dos problemas as características discriminantes são desconhecidas. É proposto um classificador de imagens similares, baseado nos pixels como características e livre da etapa de seleção, utilizando compressão de dados. Os resultados mostram que a aleatoriedade pode ser utilizada como medida de discriminabilidade, alcançando 0,83 de acurácia no conjunto de testes.**

I. Introdução

A classificação de imagens é uma tarefa desafiadora devido à grande variabilidade intraclasse ocasionada por diferentes condições de iluminação, desalinhamentos, deformações e oclusões [1]. Adicionalmente, o volume de dados multimídia tem aumentado progressivamente, potencializando aplicações de reconhecimento de padrões em imagens nas áreas de sensoriamento remoto, moda, diagnóstico médico e outras [2].

Dentre as técnicas clássicas de reconhecimento de padrões em imagens, acrescenta-se uma abordagem promissora que explora o conceito de compressão de dados para reconhecer objetos semelhantes, aproximando suas informações mútuas [3], proposta originalmente por [4]. Essa abordagem é baseada na extração de semântica por meio da entropia dos objetos, definida na Teoria da Complexidade de Kolmogorov (K). Segundo essa teoria, a entropia algorítmica de uma cadeia de caracteres é o comprimento em bits do menor programa capaz de produzir essa cadeia [5]. No conceito de aleatoriedade intrínseca de um objeto x denotada por $K(x)$, Kolmogorov estabelece um limite inferior teórico, portanto incomputável [3], para essa descrição algorítmica. Em razão da incomputabilidade de K, as pesquisas nessa linha utilizam a compressão de dados como aproximação para calcular um limite superior da complexidade algorítmica dos objetos.

Um problema clássico de similaridade de imagens é composto pelas etapas de 1) encontrar um conjunto de características que descreva de forma eficiente a assinatura das imagens e 2) aplicar uma métrica adequada para calcular as distâncias entre elas [2]. Entretanto, em grande parte dos problemas de classificação de imagens similares, são encontradas situações em que as características discriminantes são desconhecidas e/ou as fronteiras das classes não são bem definidas. Nesses casos o desempenho das técnicas clássicas é comprometido. Para [6], embora as escolhas mais populares para avaliação de similaridade de imagens ainda envolvam a extração de características, a maior dificuldade está na escolha de quais características são discriminantes para o problema. Em relação à abordagem que utiliza compressão de dados como aproximação de K, [6] afirmam que embora seja objeto de estudo há algum tempo, as pesquisas são carentes no sentido de preencher a lacuna do problema de similaridade entre imagens, sendo claramente mais eficiente para dados unidimensionais.

Nesse contexto, apresentamos os resultados parciais de uma pesquisa que propõe o desenvolvimento de um método de classificação supervisionada de imagens similares, livre da etapa de seleção de características, baseado em compressão de dados.

II. Trabalhos Relacionados

Diversos trabalhos foram propostos utilizando medidas teóricas baseadas na Teoria de Kolmogorov [3], [4]. A partir dessas formulações, foram desenvolvidas métricas aproximadas de similaridade baseadas em compressão, sendo a Distância Normalizada de Compressão (NCD - *Normalized Compression Distance*) [7], a mais amplamente utilizada. Os resultados reportados na literatura são competitivos para dados unidimensionais, porém apresentam desempenho limitado para dados bidimensionais, como no caso das imagens [6]. Isso ocorre devido à necessidade de linearizar as informações bidimensionalmente correlatas das imagens para efetuar a compressão dos dados.

No trabalho de [7] foi desenvolvida a ferramenta CompLearn, que encapsula em um pacote todo o processo de compressão de dados para clusterização. Dado um conjunto de objetos, são calculadas as distâncias entre todos os pares do conjunto. A hierarquia dos clusters é construída utilizando dendrogramas, a partir de árvores binárias com o método do quarteto. Em [8], os autores combinam técnicas utilizadas em outras áreas da ciência - NCD da Teoria da Informação, *neighbor joining* (NJ) da Filogenética e o *Fast Newman* de Redes Complexas - e testam uma nova ferramenta denominada DAMICORE. Embora o CompLearn e o DAMICORE tenham sido concebidos com o mesmo objetivo, de explorar as ferra-

mentas de compressão de propósito geral para reconhecimento de padrões, esses *toolkits* possuem diferenças na etapa de clusterização. Isso ocorre pois a NCD, por si só, não é capaz de definir agrupamentos com grandes quantidades de objetos [7].

Em [9], os autores destacam que os diversos compressores testados tiveram comportamentos bem irregulares quando aplicados em problemas de imagens. É sugerido que a NCD é diretamente dependente dos formatos de imagens e dos compressores utilizados. Portanto, a extração de semântica é influenciada tanto pela forma como os dados estão armazenados nos arquivos, quanto pelas técnicas utilizadas para compressão das redundâncias. No trabalho de [10], os autores destacam a dependência da NCD em relação à orientação da informação e que, dependendo da localização das principais características de uma imagem, os resultados podem ser inferiores. Em [11], os resultados mostram que o desempenho da NCD é ruim quando as imagens sofrem transformações geométricas. No trabalho de [12], utilizando o CompLearn, os autores discutem que os métodos de linearização linha-a-linha e coluna-a-coluna não foram eficientes no reconhecimento de similaridade das imagens rotacionadas pela NCD. Os autores ainda relatam que trabalhos anteriores obtiveram sucesso utilizando a NCD para clusterização de dados unidimensionais, diferentemente para as imagens que organizam a informação seguindo uma correlação espacial. Nesse contexto, em que a NCD se apresenta dependente dos formatos dos arquivos e técnicas de compressão e sensível à localização das informações e transformações geométricas, o método de classificação supervisionada proposto no presente artigo visa suprir, também, as seguintes desvantagens:

- Desprezo à semântica: Informações importantes de estruturas de dados bidimensionais, como as imagens, podem ser perdidas no processo de compressão dos arquivos byte-a-byte, diferentemente das estruturas unidimensionais;
- *Lazy learning*: Todo o processamento é concentrado na predição, tornando-a computacionalmente dispendiosa. A cada novo objeto submetido ao agrupamento é necessária a reconstrução da matriz de distâncias produzida pela NCD, aumentando progressivamente o volume de dados armazenados e o tempo de processamento;
- Rotulação dos clusters: O problema de rotulação dos clusters é NP-difícil e necessita de uma heurística para encontrar uma solução aproximada [13]. A rotulação pode ser comprometida pelo uso de métricas inadequadas ou falhas nas heurísticas de clusterização.

III. Metodologia

Segundo [14], a forma de percepção natural da aleatoriedade está correlacionada com a teoria matemática da Complexidade de Kolmogorov. Essa relação está diretamente ligada ao conceito de complexidade de um estímulo - no contexto do presente trabalho, o estímulo visual - ou seja, a quantidade de informações relevantes contidas em uma imagem. De acordo com [14], quanto mais complexo o estímulo, mais aleatoriedade será percebida. Portanto, quanto mais informações significantes presentes em uma imagem, maior será o valor de K, medido de forma aproximada utilizando a compressão de dados. Para preservar ou potencializar as informações relevantes de uma imagem dentro de sua classe, em detrimento das demais classes de um problema de classificação, são necessárias transformações que tanto aumentem sua complexidade intraclasse quanto simplifiquem sua descrição extraclasse. Essas transformações são realizadas utilizando técnicas de compressão de dados com perda.

Tanto a extração e seleção de características quanto a classificação são diretamente dependentes das semelhanças entre os objetos de uma mesma classe e do quanto eles se distinguem de outras classes. O presente trabalho propõe que também é possível explorar a aleatoriedade, segundo a Complexidade de Kolmogorov, desses objetos para o reconhecimento de padrões, desde que seja considerado um determinado nível de abstração para cada classe do problema. A princípio, sendo a imprevisibilidade dos acontecimentos característica fundamental do caos, não se pode afirmar que existe padrão na aleatoriedade de processos intrinsecamente indeterminados. Entretanto, os experimentos descritos a seguir apresentam razões para acreditar que é possível extrair informações e reconhecer padrões a partir da aleatoriedade intrínseca de dados correlacionados espacialmente. Vale destacar que a definição de aleatoriedade de Kolmogorov é própria de qualquer objeto que possa ser descrito algoritmicamente por uma máquina de Turing. Assim, dizer que dados correlacionados possuem aleatoriedade é diferente de dizer que são estatisticamente aleatórios.

A base de dados[1] utilizada nos experimentos foi obtida a partir do Google Imagens™ por meio de busca utilizando os nomes das classes, composta por imagens aleatórias das classes *praia*, *floresta* e *neve*. As classes foram escolhidas de modo que, na ausência de cores após todas as amostras convertidas para escala de cinza, algumas regiões das imagens, quando observadas isoladamente, pudessem ser confundidas com outras. Por exemplo, a faixa de areia de uma praia e as nuvens se assemelham à neve no chão; ou as árvores de uma floresta tropical podem serem similares às árvores de uma região montanhosa com a presença de neve; ou o mesmo céu presente nas imagens de praia pode ser observado tanto nas imagens de neve quanto de floresta. A base foi dividida em 3 conjuntos, com balanceamento entre as amostras das classes:

- Conjunto A: 33 amostras com largura de 1024 pixels;
- Conjunto B: 90 amostras com largura de 1024 pixels;
- Conjunto C: 90 amostras sem padronização de dimensões.

O *conjunto A* foi utilizado para treinamento e reúne imagens das respectivas classes sem a presença de qualquer outro elemento como pessoas, casas, carros e outros (Figura 1). Os *conjuntos B* e *C* foram destinados aos testes de viabilidade da proposta e as imagens, sempre que possível, continham outros

[1] https://www.dropbox.com/s/4hmnuc6kqmew3sp/BFSGoogle.tar.bz2

(a) (b) (c)

Fig. 1. Exemplos de imagens utilizadas para treinamento (primeira linha) e teste (segunda linha) das classes praia (a), floresta (b) e neve (c).

elementos que não foram fornecidos para o treinamento do modelo (Figura 1). O objetivo foi explorar a capacidade do método em generalizar a aprendizagem para diferentes escalas de imagens e situações com a presença de novas informações como, por exemplo, uma pessoa esquiando na neve, uma barraca dentro de uma floresta ou pessoas caminhando na praia.

As etapas do modelo são apresentadas na Figura 2. Na 1ª etapa, todas as imagens foram convertidas em escala de cinza utilizando a equação

$$I = 0,2126 \times R + 0,7152 \times G + 0,0722 \times B, \qquad (1)$$

em conformidade com a especificação ITU-R BT.709 [15], que representa a forma como olho humano percebe de maneira diferente o brilho entre as cores. Na Equação 1, o canal vermelho é representado por R e o canal azul B recebe menos peso na conversão para o nível de intensidade I (cinza) em relação ao canal verde G, em razão da visão humana perceber os tons verdes com mais brilho em comparação com os tons azuis. As 2ª e 3ª etapas são baseadas em uma técnica de compressão de imagens com perdas, chamada *quantização vetorial*.

Na 2ª etapa, as imagens de treinamento foram divididas em janelas, sem sobreposição, de tamanhos $\mathcal{J} = \{8x8, 16x16, 32x32\}$, em pixels. Em seguida, essas janelas foram linearizadas em vetores e submetidas ao algoritmo de clusterização LBG [16] por garantir distribuição mais homogênea dos centroides, de forma independente para cada classe, parametrizado para fornecer conjuntos de centroides de tamanhos $\mathcal{C} = \{64, 128, 256, 512\}$ para cada valor de $\mathcal{J}$. Os parâmetros $\mathcal{J}$ e $\mathcal{C}$ foram definidos para avaliar a manutenção e a perda de características para todas as classes do problema em diversos níveis de abstração. Assim, os experimentos foram executados em 12 rodadas com os parâmetros $\mathcal{P}$ = {[8x8, 64], [8x8, 128], ..., [32x32, 256], [32x32, 512]}, treinando e classificando as imagens com os mesmos parâmetros de janela e centroides para todas as classes.

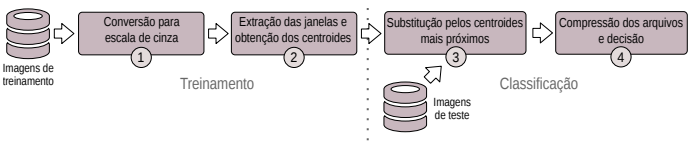

Fig. 2. Etapas do modelo proposto

Na 3ª etapa, responsável por ampliar a complexidade intraclasse e diminuir a complexidade extraclasse, os centroides obtidos na etapa de treinamento anterior foram deslinearizados e transformados em janelas quadradas. Em cada rodada, dadas uma tupla de parâmetros $\mathcal{P}$ e uma imagem de teste t, os pixels de t são substituídos pelo centroide mais próximo no espaço euclidiano, para todas as classes. Portanto, a partir de t são geradas W versões utilizando os centroides obtidos no treinamento, onde W é a quantidade de classes do problema. Por exemplo, dada uma imagem de teste t, para cada classe w, t é dividida em um conjunto de janelas $\mathcal{V}$ de tamanho j, onde $j \in \mathcal{J}$; para cada janela v de t calcula-se a distância euclidiana para os c centroides da classe w para janelas de tamanho j, onde $c \in \mathcal{C}$; o centroide mais próximo substitui v na imagem original, gerando uma versão de t aproximada de w, dita $t_{\{w,j,c\}}$.

Na 4ª etapa, calcula-se o tamanho resultante da compressão de $t_{\{w,j,c\}}$ utilizando o *bzip2* ou qualquer outra técnica de compressão sem perda. A tomada de decisão da classe predita é realizada pelo maior tamanho de arquivo comprimido. As imagens substituídas com os centroides de suas classes verdadeiras tendem a obter menor taxa de compressão quando comparadas às suas respectivas cópias substituídas com centroides de outras classes, considerando o mesmo tamanho de janela j e quantidade de centroides c.

IV. Resultados

A Tabela I apresenta os resultados obtidos pelo protótipo para cada valor de $\mathcal{P}$. Utilizando as próprias imagens de treinamento (*conjunto A*), as acurácias da predição ficaram entre 0,7576 e 0,9394. Nas imagens de teste, a acurácia mínima foi de 0,7 (*conjunto B*) e a máxima de 0,8333 (*conjunto C*), com desempenho sensivelmente superior observado nas imagens do *conjunto C*: avalia-se que o método tem desempenho invariante às dimensões das imagens, não sendo necessário nenhuma pré-processamento.

Algumas matrizes de confusão são apresentadas na Figura 3. Embora em alguns casos tenham sido observadas elevadas taxas de discriminação entre as classes *praia* e *floresta* (Figuras 3(d) e 3(f)), chegando a 1 no *conjunto A* (Figura 3(b)),

Tabela I
Acurácias obtidas pelo classificador. Com asterisco (*), os melhores resultados para cada conjunto.

Janela	Centroides	Conjunto A	Conjunto B	Conjunto C
8x8	64	0,7879	0,7	0,7333
	128	0,7576	0,7	0,7667
	256	0,7576	0,7222	0,7667
	512	0,7576	0,7111	0,7444
16x16	64	0,7576	0,7444	0,7111
	128	0,8485	0,8*	0,7889
	256	0,8788	0,8*	0,8333*
	512	0,8788	0,7778	0,7667
32x32	64	0,8182	0,6667	0,7556
	128	0,8788	0,7222	0,7444
	256	0,9394*	0,7667	0,7333
	512	0,9091	0,7	0,6444

		Predita		
		Praia	Floresta	Neve
Verdadeira	Praia	1,0000	0	0
	Floresta	0,0909	0,9091	0
	Neve	0,6364	0	0,3636

(a) *Conjunto A*, $j = 8x8$, $c = 256$

		Predita		
		Praia	Floresta	Neve
Verdadeira	Praia	1,0000	0	0
	Floresta	0	1,0000	0
	Neve	0,1818	0	0,8182

(b) *Conjunto A*, $j = 32x32$, $c = 256$

		Predita		
		Praia	Floresta	Neve
Verdadeira	Praia	1,0000	0	0
	Floresta	0,0667	0,9333	0
	Neve	0,8000	0,0333	0,1667

(c) *Conjunto B*, $j = 8x8$, $c = 64$

		Predita		
		Praia	Floresta	Neve
Verdadeira	Praia	1,0000	0	0
	Floresta	0	0,9667	0,0333
	Neve	0,5333	0,0333	0,4333

(d) *Conjunto B*, $j = 16x16$, $c = 128$

		Predita		
		Praia	Floresta	Neve
Verdadeira	Praia	0,9000	0,1000	0
	Floresta	0	0,9667	0,0333
	Neve	0,9000	0,0333	0,0667

(e) *Conjunto C*, $j = 32x32$, $c = 512$

		Predita		
		Praia	Floresta	Neve
Verdadeira	Praia	0,9333	0,0667	0
	Floresta	0	0,9667	0,0333
	Neve	0,4000	0	0,6000

(f) *Conjunto C*, $j = 16x16$, $c = 256$

Fig. 3. Matrizes de confusão com os piores (a) e melhores (b) resultados do conjunto de treinamento A, piores (c) e melhores (d) resultados do conjunto de teste B e com os piores (e) e melhores (f) resultados do conjunto de teste C.

também foi verificado baixo desempenho, devido à baixa discriminabilidade entre as classes *neve* e *praia* nos piores resultados do *conjunto A* (Figura 3(a)), assim como nos testes com os *conjuntos B* (Figura 3(c)) e *C* (Figura 3(e)). Esses resultados sugerem que a discriminabilidade do método pode ser melhorada treinando o modelo com outros tamanhos de janelas e número de centroides. Adicionalmente, acreditamos que cada classe pode ser treinada e testada em diferentes parâmetros, a fim de atender as especificidades de suas características intrínsecas. Assim, por exemplo, determinadas classes podem ser melhor discriminadas com janelas menores e maior quantidade de centroides que outras.

V. Conclusão

Nesse artigo propomos uma nova técnica de classificação de imagens similares, utilizando compressão de dados como aproximação da Complexidade de Kolmogorov, extraindo características diretamente dos pixels e livre da etapa de seleção. A principal contribuição do trabalho é demostrar que a aleatoriedade pode ser utilizada como medida de decisão para classificação de imagens. As transformações realizadas pelo método foram capazes de potencializar as características mais importantes, tornando o estímulo visual da imagem mais complexo e, portanto, mais aleatório quando substituído pelos protótipos da classe verdadeira.

Para o conjunto de treinamento a acurácia máxima foi 0,94 e para o conjunto de testes de 0,83. Os resultados indicam viabilidade da proposta, reforçando que a medida de compressão utilizada como aproximação do cálculo da complexidade de uma imagem pode ser adequada para problemas de classificação, mesmo em escalas diferentes. Entretanto, o intuito da proposta do classificador não é superar o desempenho dos métodos tradicionais ou mesmo das arquiteturas de redes neurais artificiais profundas, mas contribuir com uma abordagem livre de seleção de características. Como trabalhos futuros, pretende-se investigar se diferentes parâmetros de treinamento para cada classe são capazes de tornar mais complexo o estímulo visual e melhorar a discriminabilidade entre classes próximas; e aplicar outras medidas de aleatoriedade.

Agradecimento

Os autores agradecem ao projeto de P&D ANEEL-Copel Distribuição nº 2866-04842017 que oferece suporte financeiro à equipe de pesquisadores.

Referências

[1] T. H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.

[2] K. Juneja, A. Verma, S. Goel, and S. Goel, "A survey on recent image indexing and retrieval techniques for low-level feature extraction in cbir systems," in *2015 IEEE International Conference on Computational Intelligence and Communication Technology, CICT 2015*. IEEE, 2015, pp. 67–72.

[3] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitányi, "The similarity metric," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3250–3264, 2004.

[4] C. H. Bennett, P. Gâcs, M. Li, P. M. B. Vitányi, and W. H. Zurek, "Information distance," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1407–1423, 1998.

[5] A. N. Kolmogorov, "Three approaches to the quantitative definiion of information," pp. 3–11, 1965.

[6] A. J. Pinho and P. J. S. G. Ferreira, "Image similarity using the normalized compression distance based on finite context models," in *18th IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 1993–1996.

[7] R. Cilibrasi and P. M. B. Vitányi, "Clustering by compression," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1523–1545, 2005.

[8] A. Sanches, J. M. Cardoso, and A. C. B. Delbem, "Identifying merge-beneficial software kernels for hardware implementation," in *2011 International Conference on Reconfigurable Computing and FPGAs, ReConFig 2011*, 2011, pp. 74–79.

[9] P. P. Vázquez and J. Marco, "Using normalized compression distance for image similarity measurement: An experimental study," *Visual Computer*, vol. 28, no. 11, pp. 1063–1084, 2012.

[10] D. Coltuc, M. Datcu, and D. Coltuc, "On the use of normalized compression distances for image similarity detection," *Entropy*, vol. 20, no. 2, p. 99, 2018. [Online]. Available: http://www.mdpi.com/1099-4300/20/2/99

[11] N. Tran, "The normalized compression distance and image distinguishability," *Human Vision and Electronic Imaging XII*, vol. 6492, no. February 2007, p. 11, 2007. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.704334

[12] J. Mortensen, J. J. Wu, J. Furst, J. Rogers, and D. Raicu, "Effect of image linearization on normalized compression distance," *Signal Processing, Image Processing and Pattern Recognition*, vol. 61, pp. 106–116, 2009.

[13] R. Linden, "Técnicas de Agrupamento," *Revista de Sistemas de Informação da FSMA*, vol. 4, pp. 18–36, 2009.

[14] N. Gauvrit, F. Soler-Toscano, and H. Zenil, "Natural scene statistics mediate the perception of image complexity," *Visual Cognition*, vol. 22, no. 8, pp. 1084–1091, 2014.

[15] ITU Radiocommunication Sector, "Recommendation itu-r bt.709-6: Parameter values for the hdtv standards for production and international programme exchange," International Telecommunication Union, Tech. Rep., 2015.

[16] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, 1980.